UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/815,591	03/23/2001	Victor Spivak	EMCCP074	5572

57255 7590 09/18/2008

VAN PELT, YI & JAMES LLP AND EMC CORPORATION
10050 N. FOOTHILL BLVD.
SUITE 200
CUPERTINO, CA 95014

EXAMINER	
BASEHOAR, ADAM L	
ART UNIT	PAPER NUMBER
2178	
MAIL DATE	DELIVERY MODE
09/18/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	09/815,591	SPIVAK ET AL.	
	Examiner	Art Unit	
	ADAM L. BASEHOAR	2178	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *20 May 2008*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,4-8,10-16 and 18-21* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1, 4-8, 10-16, and 18-21* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This action is responsive to communications: The Amendment filed 05/20/08.

2. Claims 2, 3, 9, and 17 have been cancelled necessitated by Amendment.

3. All previous rejections to the claims have been withdrawn as necessitated by Amendment.

4. Claims 1, 4-8, 10-16, and 18-21 are pending in the case. Claims 1, 7, and 14 are independent claims.

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. Claims 1, 4-5, 7-8, 10-16, and 18-21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Doerre et al (US-6,446,061 09/03/02) in view of Weiser et al (US-5,982,507 11/09/99) in further view of Chakrabarti et al (US-6,418,433 07/09/02) in further view of Russell-Falla et al (US-6,675,162 01/06/04) in further view of McKeown et al (US-6,473,730 10/29/02).

-In regards to independent claims 1, 7, and 14, Dorre et al teaches a computer implemented method comprising a processor and memory connected to said processor, wherein the method further comprises:

recognizing two or more concepts in a document wherein said two or more concepts each represent a basic idea expressed in said document format (column 5, lines 57-65: "extracting for each of said unprocessed documents its features…in a feature vector");

recognizing a concept association for the two or more recognized concepts associated with a conceptual model that includes the concept association for the two or more recognized concepts (column 4, lines 65-67; column 6, lines 1-5; column 12, lines 17-67; column 13, lines 1-59: i.e. different concepts associated via the hierarchical clustered taxonomy);

indicating a concept type associated with said document using the conceptual model (column 5, lines 55-65: "unprocessed document with said category-scheme of said cluster": i.e. the document is associated with a cluster with the most similar feature-vector; column 14, lines 25-65: "categorization tool assigns documents to predefined categories"); , wherein the concept type comprises a group of one or more concepts that represent a similar idea (column 2, lines 25-45: "cluster a generalized title or cluster label…group documents by subject"; column 14, lines 26-35: "assign documents to preexisting categories, sometimes called topics or themes").

Doerre also teaches for each of said two or more concepts, identifying a plurality of features in said document format, wherein said plurality of features represent evidence of said of one or said two or more concepts in said format (column 5, lines 5-67: "extracting for each...unprocessed documents...features and computing its feature statistics in a feature-vector"; column 6, lines 1-8; column 17, lines 44-49; column 19, lines 47-65).

Doerre further teaches wherein the document could come from a multitude of documents (column 4, lines 30-32). Dorre does not specifically teach wherein the initial document format had to be converted to one of the common document format to be processed. Weiser et al teach

converting a document format (email message) from an email format to a common generic format (column 12, lines 53-55). It would have been obvious to one of ordinary skill in the art at the time of the invention for Dorre to have converted a document's initial format to a common document format, because Weiser et al taught that by doing so the common format can be understandable by the document system (column 12, lines 44-56: i.e. converting a document to a format able to be processed by the a specific system provides the obvious advantage of being able to process the document in that system).

Doerre further teaches wherein recognizing and categorizing documents was well known to enhance further document searching and retrieval (column 8, lines 37-40: "powerful and flexible queries"; column 9, lines 1-3). Doerre does not specifically teach receiving a search query associated with said concept type and identifying based at least in part on the association of the concept type with said document, that said document is responsive to said search query. Chakrabarti et al teach wherein a search query was associated with said concept type identification (column 2, lines 23-28 & 58-60: "database of Web pages that is focused on a predefined topic or topics"; column 3, lines 52-57: "focused database…receiving a search query"; column 5, lines 13-27); identifying said concept at least in part by using said concept identification of said search query (column 2, lines 58-60: "generate a database of Web pages that is focused on a predefined topic or topics"; column 5, lines 21-25: "a user can search the database 30 efficiently for Web pages of interest, i.e., only for Web pages relating to the topic of which the database 30 was focuses"); utilizing the conceptual model (column 4, lines 61-66; column 5, lines 13-27) to determine that said document was associated with said identified concept (column 2, lines 58-60; column 3, lines 52-57: column 5, lines 13-27); and concluding at

least in part on the determination that said document was associated with said identified concept, that said document was responsive to said search query (column 3, lines 52-57; column 4, lines 61-66; column 5, lines 13-27). It would have been obvious to one of ordinary skill in the art at the time of the invention for Dorre to have received a search query associated with said concept type for identifying said concept at least in part by using said identification, because Chakrabarti et al teach that utilizing models (Fig. 1: 35B & 35B) to associated documents with a predefined topic or topics (i.e. concepts) allows efficient searching of said topics by users (column 2, lines 58-60; column 3, lines 52-57).

Doerre teaches extracting features from a document and establishing a plurality of thresholds to be associated with said features (column 13, lines 24-54) by comparing said feature vectors/concept weights with a predetermined threshold (column 5, lines 28-32 & 42-67; column 6, lines 1-8; column 17, lines 44-49; column 19, lines 47-65). Doerre does not specifically teach calculating a concept weight for one of said two or more concepts using a plurality of feature weights associated with said plurality of features wherein said concept weight represents a recognition confidence level for one of said two or more concepts and comparing said concept weight with a predetermined thresholds. Russell-Falla et al teaches calculating a concept weight/confidence level for one of said two or more concepts ("calculating a rating of the page")(column 3, lines 54-57) using a plurality of feature weights ("requires a weighting be provided for each word of phrase in the list")(column 3, lines 46-57) associated with said plurality of features ("regular expressions")(column 2, lines 55-59; column 8, lines 9-19) wherein said concept weight represents a recognition confidence level for one of said two or more concepts (column 3, lines 54-57) and comparing said concept weight with a predetermined

thresholds (column 2, lines 64-67; column 3, lines 1-16). It would have been obvious to one or ordinary skill in the art at the time of the invention for Doerre to have calculated a weighting for the unprocessed documents feature values to be compared to threshold values as taught in Russell-Falla, because Russell-Falla taught that through user selectable threshold values a documents relevance could more easily tailored to a specific user (column 2, lines 64-67; column 3, lines 1-16).

Neither Doerre nor Russell-Falla teach wherein the recognition confidence level for each of said two or more concepts was calculated for each paragraph of the common document format. McKeown et al taught segmenting text documents into paragraphs, identifying the significance of specified categories of information in said paragraphs, assigning weighted scores to paragraphs of the input document, and summing the scores of the individual paragraphs for the input document (column 2, lines 35-50; column 3, lines 65-67; column 4, lines 4-15, 29-34, & 45-57; column 5, lines 9-45; column 6, lines 21-30 & 53-57; column 7, lines 1-15 & 38-63). It would have been obvious to one of ordinary skill in the art at the time of the invention for the confidence level scoring of documents for specific concepts of Doerre and Russell-Falla to have been calculated on the paragraph level of said documents as shown in McKeown et al, because McKeown et al taught that segmenting document scores by paragraphs provided the benefit of "efficiently and accurately" identifying topics in an input document (column 2, lines 30-34) which provided the additional well known benefit of helping text search engines process input queries to locate specific topics in a given document (column 1, lines 40-50: "be useful in assisting text search…useful in assisting summary agents").

-In regard to dependent claims 4 and 19, Doerre teaches based on conceptual model, generating an auto-attribute, said auto-attribute being a descriptive label for said document (column 2, lines 25-45: “cluster a generalized title or cluster label…group documents by subject"; column 14, lines 26-35: "assign documents to preexisting categories, sometimes called topics or themes").

-In regard to dependent claims 5, 18, and 20, Doerre teaches based on said conceptual model assigning said common format document to a subject category in a directory (column 13, lines 1-24; column 14, lines 26-65).

-In regard to dependent claim 8, Dorre teaches wherein said conceptual model includes a concept dictionary (column 14, lines 57-65: “category scheme is a dictionary”).

-In regard to dependent claim 11, Doerre teaches wherein the conceptual model includes a noise dictionary (column 6, lines 37-47).

-In regard to dependent claim 12, Doerre teaches assigning a subject category to said document based at least in part upon said conceptual model (column 12, lines 61-67; column 13, lines 1-54).

-In regard to dependent claim 13, Doerre teaches wherein assigning the subject category follows an auto-categorization rule (column 5, lines 55-67; column 6, lines 1-3; column 14, lines 57-67; column 15, lines 1-20).

-In regard to dependent claim 15, as shown above, Doerre teaches wherein the document could come from a multitude of documents over the Internet (column 1, lines 18-24; column 4, lines 30-32). Doerre does not specifically teach wherein the initial document format had to be converted to one of the common document format to be processed. Weiser et al teach converting a document format (email message) from an email format to a common generic format (column 12, lines 53-55). It would have been obvious to one of ordinary skill in the art at the time of the invention for Dorre to have converted a document's initial format to a common document format, because Weiser et al taught that by doing so the common format can be understandable by the document system (column 12, lines 44-56: i.e. converting a document to a format able to be processed by the a specific system provides the obvious advantage of being able to process the document in that system).

-In regard to dependent claim 16, Doerre teaches separating the text content from said initial format document for categorizing documents based on text analysis techniques (column 5, lines 55-65: "extracting for each of said unprocessed documents its features"; column 9, lines 5-67; column 10, lines 1-67; column 11, lines 1-61). As shown above in dependent claim 15, Doerre does not teach converting the initial document format into a common document format. Weiser et al teach converting a document format (email message) from an email format to a

common generic format (column 12, lines 53-55). It would have been obvious to one of ordinary skill in the art at the time of the invention for Doerre to have converted its initial format document to one of the common document formats listed above, because Weiser et teach by doing so the common format can be understandable by the document system (column 12, lines 44-56)(i.e. converting document to a format able to be processed by the a specific system provides the obvious advantage of being able to process the document in that system), wherein it would have also been obvious to incorporate the text from the initial document into the said common document, because Doerre teaches the textual content was what was needed to categorize the incoming documents (column 9: Section "Text Analysis Functions").

-In regard to dependent claim 21, Doerre teaches wherein the conceptual model includes a concept association dictionary (column 13, lines 1-24: "tree constructed…contains the complete clustering information including all inter- and intra-cluster similarities").

-In regard to dependent claim 10, Doerre teaches setting specific threshold values for concept inclusion into the conceptual model at a plurality levels based on exceeding and or meeting certain thresholds (column 13, lines 1-55: "set of singleton clusters each containing a single documents…allows the user to set…thresholds of intra-cluster similarity")

7. Claim 6 is rejected under 35 U.S.C. 103(a) as being unpatentable over Doerre et al (US-6,446,061 09/03/02) in view of Weiser et al (US-5,982,507 11/09/99) in view of in further view of Chakrabarti et al (US-6,418,433 07/09/02) in further view of Russell-Falla et al (US-

6,675,162 01/06/04) in further view of McKeown et al (US-6,473,730 10/29/02) in further view of W3C's, "Extensible Markup Language (XML) 1.0", 02/10/98, pp. 1-2, http://www.w3.org/TR/1998/REC-xml-19980210.

-In regard to dependent claim 6, Doerre teaches wherein the document could come from a multitude of documents over the Internet (column 1, lines 18-25; column 4, lines 30-32). Doerre does not specifically teach wherein a common format was an XML document. W3C teaches wherein using XML was notoriously well known in the art for web applications (pp. 2: Section 1.1). It would have been obvious to one of ordinary skill in the art at the time of the invention, for one of the common formats of Doerre to have been XML, because W3C teaches that the XML format provides the benefits of being easy to create, being easy to write programs which process XML documents, and being human-legible and reasonably clear (pp. 2: Section 1.1). It was also notoriously well known in the art at the time of the invention that XML was an International document standard.

Response to Arguments

8. Applicant's arguments with respect to claims 1, 7, and 14 have been considered but are moot in view of the new ground(s) of rejection.

Applicant's arguments with respect to the Russell-Falla reference fail to comply with 37 CFR 1.111(b) because they amount to a general allegation that the claims define a patentable invention without specifically pointing out how the language of the claims patentably distinguishes them from the references.

Conclusion

9. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

10. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Please note the additionally cited prior art on the accompanying PTO-892 form.

11. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Adam L. Basehoar whose telephone number is (571)-272-4121. The examiner can normally be reached on M-F: 7:00am - 4:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Steve Hong can be reached on (571) 272-4124. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/Adam L Basehoar/
Primary Examiner, Art Unit 2178